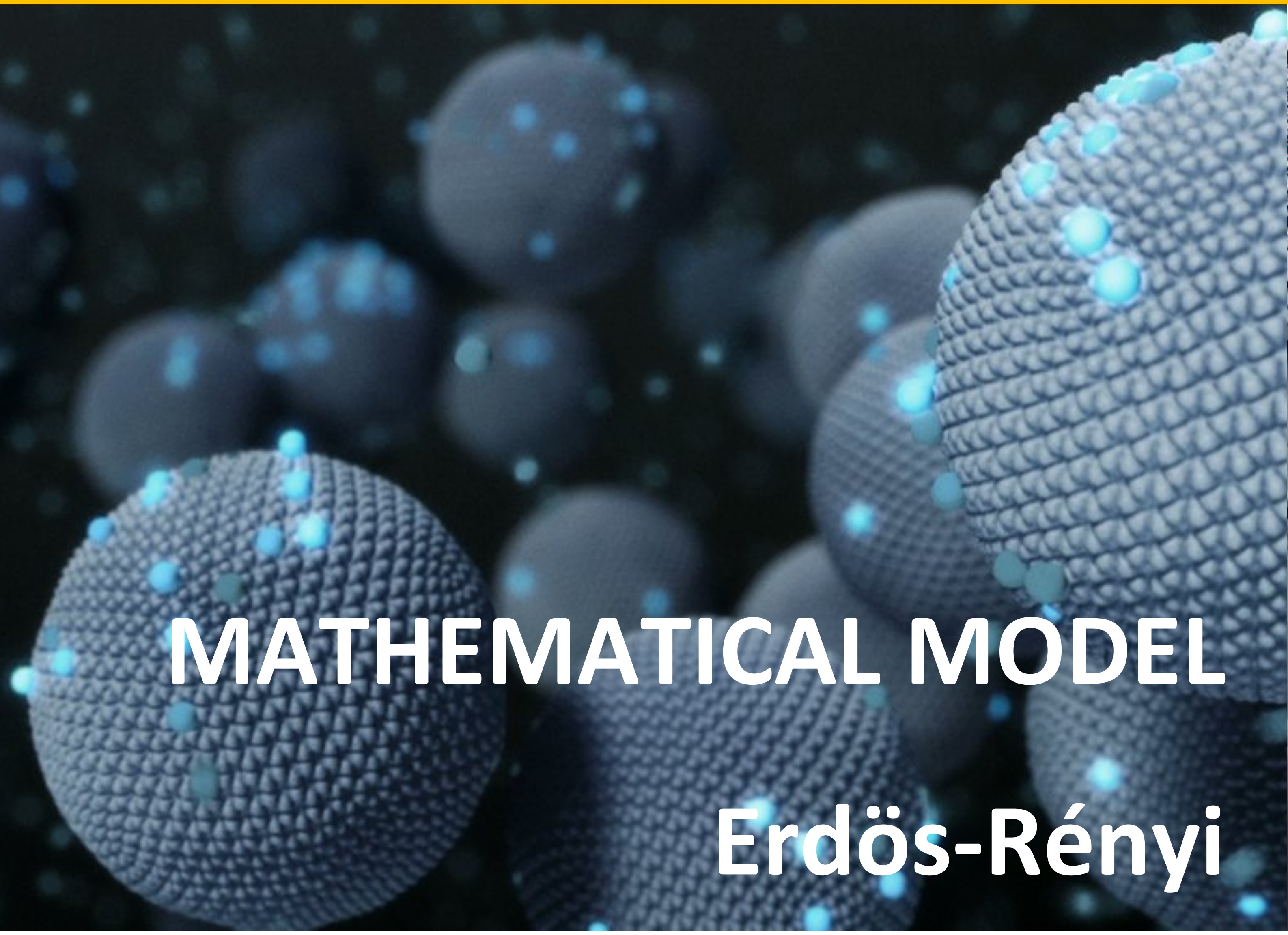

MATHEMATICAL MODEL

Erdös-Rényi

BIO IQS
iGEM · BARCELONA

Prepared by
bioiqs.igem@gmail.com
Laboratory of Biochemistry, IQS School of Engineering

IQS PERSONA CIENCIA EMPRESA UNIVERSITAT RAMON LLULL

A IQS

Biokit A Werfen Company

GUIVES 1968

Erdos-Renyi model

Under the considerations of space free of spatial correlations within a set of connected or not elements under mean field assumptions. Given a graph such Ω = (V,L) defined by a set of N vertices or nodes and a set of L edges or links (R.Solé, Phase Transitions; 2011):

$V=\{v_1, v_2, v_3,\dots,v_n\}$ $E=\{e_1, e_2, e_3,\dots,e_n\}$

It is possible to define k_i as the number of edges connecting v_i with other nodes in the graph. Subsequently, the average degree $\langle k \rangle$ is:

$$z = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i$$

Under the assumption of every edge or connection between two random nodes exist with a probability *p*. then the average number of edges $\langle L \rangle$ in the graph is:

$$\langle L \rangle = \frac{N(N-1)}{2} p$$

Where N is the number of nodes, so N(N-1) takes in account all the possibilities of node couples connected by an edge at probability *p*, divided by 2 because each connection requires a pair of nodes. Then retreating to average degree concept at large N:

$$z = \frac{z\,\langle L \rangle}{N} \approx Np$$

It can be easily shown, at large quantity of nodes, the average in the number of nodes connecting each node to the others I directly related to the quantity of nodes and the probability of the link presence between them. Then, the probability that a random node has a degree k, and under the assumption that one node could have some edges or null, a binomial probabilistic equation exists:

$$P(k) = \frac{(N-1)!}{K!\,(N-1-K)!} \cdot p^k \cdot (1-p)^{N-k-1}$$

The first term takes in account all the existing combinations between nodes and their possible k degrees, k element taken from N in k. The second term aggregates the probability of have k nodes, and the third term the probability of having none. In statistics a binomial distribution is a specific probability distribution that takes in account the number of success (k degree in a random node), in a sequence of N Bernoulli assays, independent between them. Following a fixed probability of success occurrence among the assays.

A Bernoulli assay is a random experiment where only 2 results are possible. In our case, a k degree in a specific node, a determined number of links emerging from a single random node or in opposition, not a k degree in the same node. Simplified, success or failure. The for a large N and fixed z, a binomial Poisson distribution fits well, being the probability distribution of the number of success in a sequence of N independent experiments at probabilities $p_1,p_2,\dots,p_n$. If z = N*p*:

$$P(k) = e^{-z}\frac{z^k}{k!}$$

Where a randomly chosen node is likely to have a degree z or $\langle k \rangle$ or close to it.

No there are described mathematically the probability of having z number of links at a probability *p*. If p is small the probability that two random nodes are connected by some path is minimal and the system will be fragmented in small clusters of little number of nodes interconnected, but when *p* is large enough a great number of nodes will be connected by some path.

Subsequently, as we increase z, how varies the probability of finding two random nodes interconnected by one or more paths. In other words, the probability of observe a more connected network increases with z? how is this increase?

An interesting concept rises from the question. The giant component of the network is the largest cluster of the network and could involve only few nodes to all of them, depending of the connectivity of the network (Bollobas 1985). In order to understand the dynamics underlying the formation of the giant component at different z values, it must be understood how the giant component G(*p*) varies as *p* changes.

Considering two groups in the network, the group that collects all the nodes belonging to the giant component G_∞, and who's not Q. The giant component plus the rest forms the hole, G=1-Q. then the probability of not belonging to the giant component is: Q = P [$v_i \notin G_\infty$]. Then the key assumption is that the probability that a random node of k edges is not belonging to the giant component must be equal to the probability of any of its k neighbors is not part of the giant component:

$$Q = \sum_{k=0}^{\infty} P(k) Q^k$$

From the previous equation of P(k) probability:

$$Q = \sum_{k=0}^{\infty} e^{-z} \frac{z^k}{k!} Q^k$$

$$Q = e^{-z} \sum_{k=0}^{\infty} \frac{(zQ)^k}{k!}$$

Following a Tailor expansion $\sum_{y=0}^{\infty} \frac{(x)^y}{y!} = e^x$:

$$Q = e^{-z} e^{zQ} = e^{z(Q-1)}$$

Finally

$$G = 1 - Q = 1 - e^{z(Q-1)}$$

$$\boldsymbol{G = 1 - e^{-z(G)}}$$

The graphic representation of the equation:

$$z = \frac{\ln(1-G)}{-G}$$

$$\begin{cases} G = 0 \; if \; z < 1 \\ G \neq 0 \; if \; z > 1 \end{cases}$$

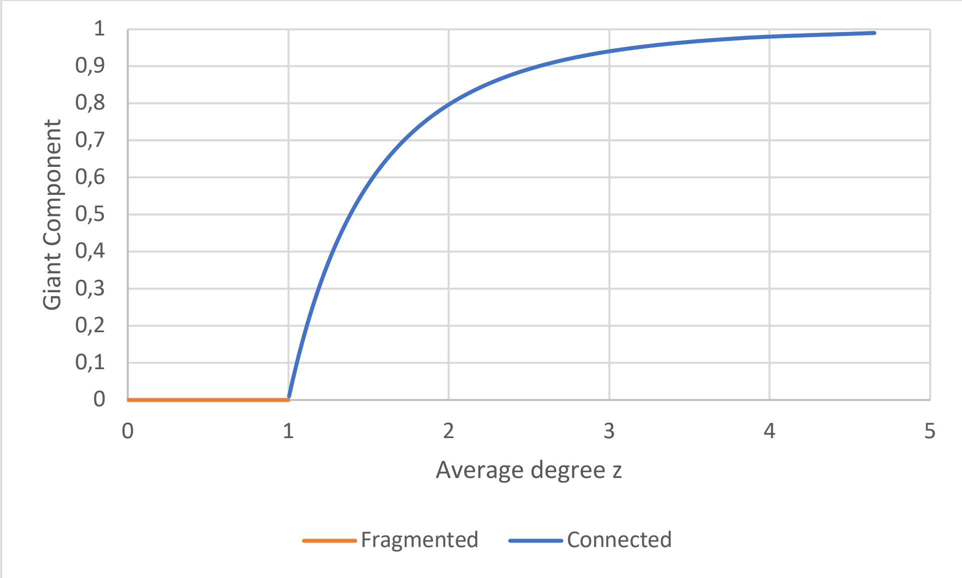

Figure 1. The fraction S of the graph belonging to the giant component, while the average degree z is tuned. As in can be easily seen, a phase transition occurs just at $z_c = 1$. The stable state of the system drift from fragmented networks to a connected hole network.

The bifurcation diagram exposed above summarizes self-explicative information about the process of network formation. The size (probabilistic speaking) of the greatest component in the system is represented by the y axis, following 0 to 1 scale being 0 no existing any giant component and 1 when the network is fully connected. At the x axis, the average degree z is represented, as the number of edges connecting a random chosen node with other nodes in the graph. The representation explores the average probability of finding the network in a specific state. Collective properties arise at critical point.

The most dramatic characteristic of those systems is the breaking of symmetry that happens at z_c. The change in the dynamics is abrupt as we tune up the value of z from 0 to ∞. The probability of finding a network fully connected doesn't increase at a constant ratio, there exists two states well delimited between a connected network and a fragmented one. In one hand, an average of links starting from a chosen node inferior to one, the resulting scenario will be from fully disconnected beads at maxim entropy to little clusters of beads with none predominant, depending on the ratio links/nodes. However, an average of links starting from a chosen node superior to one, will lead, at certain probability to a connected network. The shift is abrupt, and the formation of a better network doesn't follow a linear tendency.

In addition, it's interesting to notice that as z increases to $z \rightarrow \infty$, asymptotic approximation to a fully connected network occurs. Consequently, increasing $z \rightarrow \infty$ doesn't results in a linear increasing network, providing the value idea that a low number, but enough, connections per node (average) could be enough to acquire a sufficiently connected network.

It is true that low number of links in relation to the number of nodes are enough in order to create a connected network, and mort importantly, when the criticality is reached, the connectivity arose at minimum costs. This is a powerful idea, taking in account that the number of nodes and the number of linkers and the HLA necessaries to connect them are limited resources and exist a great interest in their minimization, directly related to the cost minimization.

Another important feature of those systems in direct application to the UV-vis sensor, is that accuracy primary specifications, could be primarily addressed. Under the assumption that a hypothetic scenario of little amount of traces present in the food sample, it will be important to manage a censoring methodology capable of detecting such difference. Great amount of HLA per node, or equally great number of potential linkers per node will lead the system to a connected state, even if a part of the connections are blocked. Tuning the amount of HLA over the surface into a number close to the criticality, would lead the system into a tight detection window.